Facebook graph search for refined screen-based data collection in CMC: A pilot study for Fiji English

Huber, André

DOI: https://doi.org/10.1016/j.dcm.2015.03.001

Facebook Graph Search for Refined Screen-Based Data Collection in CMC – A Pilot Study for Fiji English

André Huber
University of Zürich, English Department, Plattenstrasse 47, 8032 Zürich, Switzerland.
andre.huber@es.uzh.ch, (+41) 44 634 39 37

Abstract
In sociolinguistics, the advantages of web-based research are often lost because it can rarely satisfy the need for social variables without additional offline data. Facebook Graph Search is a new query tool, which allows highly specific retrieval of Facebook users and can thus mitigate this dilemma. In this paper, Facebook Graph Search will be introduced and a pilot study in computer-mediated communication and Fiji English will demonstrate the suitability of Graph Search for exploratory research in a speech community. The pilot study shows that there are differences in language choice based on age and context of use: English is preferred over local indigenous languages by younger users and in wall posts. The data further revealed not only the use of typical CMC features, but also locally influenced features and spellings.

Keywords
Computer-mediated Communication; Facebook; Graph Search; Methodology; Fiji English

1. Introduction

"The internet constitutes a 'field site' with enormous potential for sociolinguists, who have at their disposal a wide range of language use, which can be studied using different methodologies and to answer various research questions." (Bolander and Locher 2014: 18) Despite the sheer amount of available data, the seeming ease of access compared to cost- and time-intensive sociolinguistic fieldwork, and the possibility to avoid the observer's paradox, embracing accounts of the internet as data source are usually followed by reservations. Many language samples taken from the internet are still guised in anonymity or at least not fully transparent regarding their creators' identities. Especially in sociolinguistics this is a serious problem. Research on computer-mediated communication (CMC) has usually addressed this problem by limiting the observed online data to people known in the offline world. Androutsopoulos (2007b), for example, conducted interviews with webmasters and members of an online community he observed beforehand; thus, combining online and offline data. With the evolution of the internet, more and more research on Facebook has emerged in the last few years and also there this problem has been approached in different ways. Some researchers focused on Facebook groups, in which users can be considered homogenous in terms of the particular interest covered by the group (e.g. Honeycutt and Cunliffe 2010), others looked at Facebook pages, i.e. public profiles of institutions, companies, and public figures (e.g. Steinfeld and Lev-On 2014), but most collected wall data (i.e. status updates, posts, comments) from people in their personal network (e.g. West 2013) or from people willing to participate in the study (e.g. Sultana 2014). The last approach is obviously the most laborious way of collecting data, as the advantage of instant data availability in the web context is lost. The reason most researchers settle for this option anyway is not only the need for more transparent user data, but also to get access to private profiles and content, or to get participants' consent for using their public data. However, a significant portion of linguistic research often consists of "systematic observation" (Androutsopoulos 2008), i.e. taking a first glance at relevant data in order to build hypotheses, or even test existing hypotheses in a first environment, where considerations such as consent for publication are not of relevance yet. Androutsopoulos calls this "screen-based" data collection (2013: 240). Facebook seems like the ideal platform for such a purpose as people from all around the world use it on a regular basis. The only problem so far was the retrieval of those language users.

This paper will first introduce a new query format available on Facebook called "Graph Search" that greatly facilitates the retrieval of relevant users. It thus provides an approach to mitigate the problem of "the relative anonymity of many online environments for the elicitation of core information on social variables like age, gender and social class background" (Bolander and Locher 2014: 24). Then, a pilot study for CMC in Fiji will demonstrate how Facebook Graph Search can be used to obtain a first insight into a linguistic community right from the researcher's desk without having to invest much time and money. As there is hardly any existing literature about CMC in Fiji, it is a prime example for such an undertaking.

2. Facebook Graph Search

In order to illustrate how Facebook Graph Search works, I will give a generic example: To find people on Facebook from Zürich, Switzerland who speak Swiss German, the only alternative to established contacts and offline networks has been to either search for groups that likely attract people with the desired traits, or for researchers with programming skills to use the Facebook API[1], which already allows to retrieve users based on location (if provided). But rough search criteria as these are mostly not precise enough for linguistic research. A case in point would be a person providing Zürich as his or her residence who just moved there. Obviously, more information would be helpful and many Facebook users are apparently willing to provide it. But filtering through all the members of a Zürich-interest group to check for further information like languages spoken, age, place of residence or whatever else might be relevant for a given research interest is a time-consuming task. This is where Facebook Graph Search comes in. Graph Search is not exactly brand new, it was introduced in January of 2013, but it is still a beta version and currently only available to people who use the platform's US-English language interface[2]. In a nutshell, Graph Search can be explained as follows:

> Basically, it's a revamped search tool that lets you find the friendly needle in your social media haystack by looking through all Facebook content that's been shared with you or is public - that includes friends, friends' interests, photos, games, apps and so on. (Solomon 2013)

The actual clue, however, is the way you can get to that needle – namely, by combining all the desired traits and interests into one query. Accordingly, it becomes possible to search for people who live in Zürich *and* were born in Zürich *and* speak Swiss German *and* are male, etc. The possibilities for combinations are almost endless and can also include things a person likes (for example a certain type of music, a museum, hobbies, etc.). The use of this feature is very obvious for marketing purposes (i.e. getting to know one's potential customers), but for the average user the function probably does not exactly serve a want. As a researcher, however, Graph Search can be very useful, as this paper is meant to show.

Graph Search functions quite intuitively by entering a search string into the normal query field. The query-syntax allows formulating a command in a natural language-like manner, i.e. by coordinating multiple parameters with *and*, specifying parameters with

---

[1] API stands for "application programming interface" and provides predefined means for programmers to access Facebook, for example when implementing a cell phone app that needs to access a user's Facebook account. https://developers.facebook.com/docs/graph-api

[2] Given the often short lifespan of technological innovations, it comes as no surprise that since the completion of this paper changes have been made to Graph Search. In December 2014 the service was extended to the mobile interface (US-English users only) and a keyword search was added. While this is not discussed in this paper and some specific elements of the here described query function may become obsolete at any time, it is the undoubted future importance of semantic search that makes the pioneering discussion of such an example service for linguistic research important and worthwhile nevertheless.

defining relative clauses, and restricting results with time indicators. The algorithm understands the core vocabulary that indicates the relationship between the searched category and a given parameter, most prominently *like* for anything which a user indicated to appreciate by clicking Facebook's famous "like" button, or for example *speak* for languages the user provided in the respective field in the personal profile. The algorithm automatically detects which parameters are searchable, so called "nodes" in the Facebook "Graph", and then highlights them in the query string. As one types, Graph Search suggests ways of completing or alternating the query. A sample query can be seen in figure 1.

Figure 1: Query to find people from Zürich who speak a language starting with "Swiss-". Facebook's suggestions include the intended "Swiss German".

If queried as in figure 1, Facebook retrieves a list of people from Zürich and then displays a subsample of all those who speak Swiss German. Facebook will indicate a rough number of results, which does not correspond to the actual number of displayed results, however. This discrepancy seems to stem from the fact that the number mirrors what Facebook finds in its system, while the displayed results are (rightfully so) limited to those which are public or accessible due to friend relations. How powerful Graph Search is becomes even more apparent when looking at the filtering options. The results can be manually filtered by the boxes provided, or by further specifying the query string. A query for people can be further filtered by dozens of parameters, covering mostly biographical information that users can provide, such as school, employer, relations to other users, etc. Figure 2 shows a screenshot that contains all of the above: (parts of) the result list, the number of results (here "More Than 1,000 People"), and (parts of) the filter options.

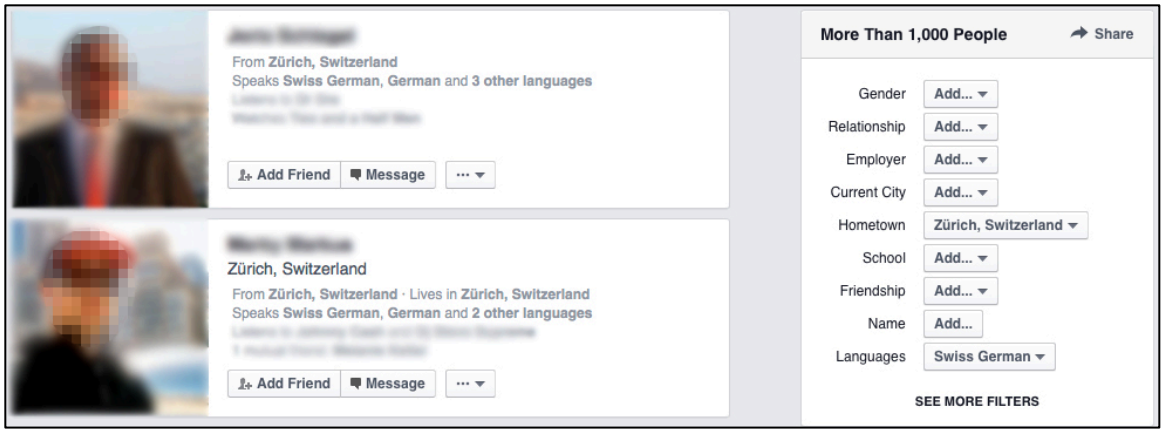

Figure 2: Query results from a search for people (left), Facebook's rough result count (top right), and manual filter options (bottom right)

It is fully understandable that many people will react with a feeling of unease about the straightforwardness of retrieving all this information. However, privacy per se is not directly affected by this new feature, which means that Graph Search displays the same amount of information a user would be able to view otherwise, all depending on the personal settings of each user. The only real "change" is the kind of privacy that previously resulted from being a needle in a haystack, whereas now the haystack no longer provides much covering. Therefore, any ethical considerations a researcher working with Facebook had to make so far still fully apply when working with Graph Search. Facebook has taken measures of their own to prevent minors from exposure to the public gaze: researchers will not be able to retrieve profiles of minors unknown to them. Whereas this protects the probably most vulnerable group, it is up to the researcher to be aware of the fact that publicly available profiles may not only result from a user's willingness to share everything with the world, but for example from ignorance about certain options in the network's sometimes complex security settings. Therefore, it is imperative to choose wisely what kind of information can be shared with the research community.

Besides the age barrier, there are additional limitations to what Graph Search can do. One major drawback is the absence of a negation operator in the query syntax[3]. Similarly, a Boolean *or*-operator is missing, slightly limiting the retrieval for two distinct sets in one go. Given that Graph Search is still a beta version only available through the standard interface and not as API for programming, negation and exclusion still necessitates manual filtering.

While most discussed limitations of Graph Search are inherent problems specific to Facebook Graph Search, there are some aspects that may apply to any comparable technology and need to be pointed out. Although privacy has been addressed above, it leads to related issues such as recall problems. A query relying on public data is producing false negatives, which is a minor issue for the kind of research proposed here, but skews the results. Also, the search algorithm, although described as a straightforward overlapping and subsampling of sets above, is almost certain to contain more

[3] Note that there is one specific case in which *not* is understood, namely to exclude one's friends from a query, e.g. "people who are not my friends and like X"

sophisticated and personalized elements that create issues of transparency and replicability. This kind of data retrieval cannot completely eliminate possible manual post filtering. The manual precision and plausibility test I carried out in the pilot study below, for example, may have to be adapted for any new research context and the resulting time and labor may vary accordingly.

3. Fiji Pilot Study

Fiji English is one of the best-researched ESL varieties in the South Pacific (Zipp 2014, 194); however, there is no research based on recent enough data to account in its analysis for CMC influences. The rise of CMC is a more current phenomenon in the Pacific Islands than it is for example in Europe, where computers have been widely available for over two decades. In Fiji, the digital revolution lagged behind for two main reasons: computers and mobile connectivity were too expensive for most people to afford. The first problem mostly disappeared with the introduction of the more affordable smartphones, while the second one was due to a service provider monopoly, which ended in 2008. Ever since, Fiji has become a CMC forerunner in the Pacific Islands in many ways, for example by having the highest Facebook population with 200,000 accounts in 2012 (Cave 2012: 9).

The sociolinguistic situation in Fiji is rather complex: the majority of people are of one of two ethnicities – indigenous Fijians and Indo-Fijians – and speak their own languages – Fijian and Fiji Hindi – while English is spoken at varying degrees of competence as a second language. Indo-Fijians are the descendants of indentured laborers that were brought to the islands between 1879 and 1916 to work on sugarcane plantations and made up over half of the population in the past – a number which has decreased to roughly one third according to the latest census from 2007. English was introduced by the British colonizers as a bridging language between the two communities and today is said to prevail "in most official spheres" such as administration, education and media (Tent and Mugler 2008, 235). Information and predictions about the use of English in Fiji in the context of the internet and the new media are limited to mostly side remarks in research articles with different main foci.

In order to get a detailed picture of the current language practices in the CMC context in Fiji fieldwork is almost inevitable. But access to remote language communities (from the perspective of a European researcher) is costly in terms of both money and time. Therefore, it is important to arrive on site with a primary idea of what to expect in order to reduce time needed for orientation. When previous research is as sparse as in the given case, this becomes even more important. With Facebook Graph Search it is now easier than ever to do exactly this and in what follows I will present a first glance into language use in the CMC context in Fiji. For this purpose, the following questions were of interest: What do people in Fiji use Facebook for? Which languages do they use on Facebook? Is there something like a CMC language, possibly even specific to the Fijian context?

4. Methodology

For the purpose of this pilot study I collected a total of 400 Facebook wall posts from 40 different users, sampled in order to get a distribution in age and ethnicity. Accordingly,

the data sample contains 200 wall posts from 20 indigenous Fijians and 200 wall posts from 20 Indo-Fijians. Both subsamples can be broken down into 100 wall posts of users who are 18-20 years old and 100 wall posts of users 25 and older. These age categories have been chosen as a result of two constraints: the Graph Search age barrier that excludes minors, and scarcity of data. While it was easy to retrieve enough people in the comparatively narrow span of early adulthood, forming more narrow categories for "older" users would have been impossible. This can be interpreted in terms of baseline frequency, i.e. there simply may be more younger Facebook users, or as a behavioral pattern, i.e. younger people may more readily provide age information on their profile. The gap of five years between the two categories is an attempt to make the two groups more distinct.

The following queries where used to retrieve the first two subsets of users:

(i) People under 21 years old from Fiji who live in Fiji and speak Hindi

(ii) People over 24 years old from Fiji who live in Fiji and speak Hindi

Queries (i) and (ii) only differ in terms of age, but contain two important assumptions made to retrieve the desired group of people. First, the location parameter Fiji is asked for twice, namely not only as place of current residence, but also as provided hometown. This was assumed to mostly rule out users who had only moved to Fiji later in life and was meant to corroborate native speakers status in each group. Second, the language parameter was used as a key to the ethnicity variable. Although there are indigenous Fijians with some competence in Hindi, the correlation of language and ethnicity is quite strong in this case. Surveys from Tent (2009) suggest that the opposite situation, i.e. Indo-Fijians using Fijian, is much more likely and was taken into account for the remaining queries as explained below. From the resulting profiles, the first 10 suitable ones were selected. Suitability in this case meant on the one hand passing a final shallow test of result precision, by simply checking whether the profiles are plausible candidates based on photos and family names, and on the other hand that the wall actually contained at least 10 posts. Wall content was chosen if a post was actively posted by the user, i.e. regular wall posts (status updates) or content shared on one's wall from different sources. Posts by other users were neglected and so was semi-automatically generated content such as reports about changes to the profile picture or the addition of pictures to an album, which both result in wall posts. The wall posts were captured via print screen in order to ensure lasting availability. This is another area where Graph Search access via API would facilitate data collection. Posts were captured including comments; however, only comments made by the profile owner were investigated in order not to render the carefully constructed samples void.

The remaining subsets of users were retrieved with the following queries:

(iii) People under 21 years old from Fiji who live in Fiji and speak Fijian

(iv) People over 24 years old from Fiji who live in Fiji and speak Fijian

For queries (iii) and (iv) the same assumptions apply as for queries (i) and (ii) above, but having observed that this query did yield users with family names suggesting Indian heritage and Graph Search not providing a negator in order to find people who speak Fijian but *not* Hindi, I manually filtered the results for users who did not list Hindi as additional language. Again, after passing the same shallow plausibility test, the first 10

suitable profiles were selected. Although gender was not controlled for in the sampling, all four subsamples are quite balanced in terms of gender distribution.

Despite an even distribution regarding age, ethnicity and gender, there is an inherent skew towards Facebook users with a public profile. However, this only concerns the accessibility dimension (public/non-public) of what Landert and Jucker (2011: 1425) propose to be a multi-dimension variable, and not content (non-private/private). Further unknown variables are social class and whether the users live in an urban or rural environment. Although both were neglected in order to counteract data sparseness, either variable could in theory be accounted for in Graph Search.

The data was then coded manually for content and language use by the author. Coding decisions and category choices are addressed in the discussion below.

5. Results

5.1 How Facebook is used

The first interesting discovery concerns the use that people in Fiji make of Facebook. To be precise, the data shows what people post on their wall. The first categorization in table 1 reveals of what type the 400 collected wall posts are. Interestingly, posts to one's own wall were not as frequent as one would expect. Purely based on impression during data collection, the most space on many users' walls seemed to be taken up by frequent updates of the profile picture (i.e. a post showing the picture with the title "User X updated his/her profile picture") and the addition of photos to a photo album (i.e. a post showing one or multiple pictures with the title "User X added a photo/photos to the album X").

Table 1: Type of wall posts

	Fijians 18-20 years	Fijians >= 25 years	Indo-Fijians 18-20 years	Indo-Fijians >= 25 years	Total
Text only	59	37	50	39	185
Picture only	1	5	10	17	33
Picture and Text	22	16	27	16	81
Video only	0	0	0	1	1
Video and Text	5	5	3	0	13
Shared Picture	4	20	7	16	47
Shared Video	3	2	1	2	8
Shared Status Update (Text)	0	6	1	2	9
App Results	4	4	1	0	9
Miscellaneous	2	5	0	7	14

The categorization in table 1 reveals that overall most communication happens in writing and that even most pictures are accompanied by text, usually a comment or explanation about what the picture shows. Still, visual content must be acknowledged as a significant element in Facebook communication. As "picture" counted anything posted in the form of a picture file, i.e. not only photos, but also pictures containing text, for example a

quote. Accordingly, for a post to fall under the category "picture and text" there needed to be text outside the picture, i.e. writing from the user. The category "app results" refers to posts that display results of entertaining online tests and quizzes, such as "What animal are you?", the results of which are usually shared on the wall, often with an added comment about the outcome. Rare formats that did not fit any other categories were grouped under "miscellaneous". More interesting than just the overall numbers are the distributions among the subgroups. There it can be seen that the younger subgroup of both ethnicities produces more text (both in the form of "text only" and "pictures and text"), a difference which the older group compensates by sharing more content.

Whether more sharing also entails more interaction, or whether this is simply a different form of interaction may become clearer when taking additional information into consideration. One common way to measure interaction on Facebook is the so-called "Engagement Rate"[4]. As the amount of friends is not visible for all users and this measure gives equal weight to likes and comments, i.e. clearly distinct levels of interaction, a more appropriate measure is needed. I consider likes and shares to be negligible acts of interaction, as only comments are verbally communicated reactions to a post that can be viewed as an actual dialogue. In addition, rather than just counting the number of comments, it appears relevant to take into consideration how much users contribute to that number themselves, as well as how many unique users contribute to the comments. Although baseline frequency, i.e. the total number of friends, may influence both the amount of comments per post as well as the amount of unique respondents, the unavailability of this figure should not be a significant problem, as previous research has shown that "interaction activity on Facebook is significantly skewed towards a small portion of each user's social links [i.e. friends]" (Wilson et al. 2009: 217). Therefore, what appears both accessible and relevant for user interaction is summarized in table 2 for each subgroup.

[4] Facebook does not officially provide the formula behind the numbers they offer to page owners for statistical insights, but it appears to be common knowledge in blogs and discussion threads that the basic formula for Engagement Rate is: Total Engagement (Likes + Comments + Shares) x 100 / Total Friends.

Table 2: Comment statistics per subgroup (group averages)

	Fijians 18-20 years	Fijians >= 25 years	Indo-Fijians 18-20 years	Indo-Fijians >= 25 years
Comments per post (average)	3.29	1.56	7.7	3.87
Posts with comments (average)	72.0 %	34.0%	77.0%	55.0%
Comments per post (If-average)	4.56	3.37	8.95	5.66
Std. Dev. If-average	(2.09)	(3.37)	(6.01)	(5.73)
Self-Comments per post (If-average)	1.30	1.80	2.18	2.02
Unique-user rate (If-average)	84.3 %	93.0%	74.6%	87.0%

If-Average = Average of comments if only posts with comments are considered; Unique-user rate = Percentage of comments made by different users, including post owners

The numbers in table 2 reveal that, on average, a Facebook post draws anything between 1.56 and 7.7 comments, depending on the subgroup. However, this ignores the fact that some posts will not be commented on at all. In my data, an average of only 34% of all posts were commented in the lowest group, and 77% in the highest group. Therefore, the average of comments per post rises if only those with comments are regarded (If-average), then ranging from 3.37 to 8.95. The calculated standard deviation for the If-average shows that the Fijian subsample is much more constant in its pattern, whereas the Indo-Fijian subsample shows much more variation among individual users. Also, it is of interest how much users contribute to comments themselves, which is on average anything between 1.3 and 2.18 comments per post with comments. Finally, comments can either be an interaction between a few people or many contributions by different people. The unique-user rate reflects how many individual users contribute to the comments, i.e. 100% means that each comment came from a different user. From these figures it can be concluded that for either ethnicity the younger subgroup not only gets its posts commented more frequently, but also gets more comments per post, which in turn come from fewer unique users, suggesting more dialogue-type interaction. The higher numbers for the Indo-Fijian subgroup regardless of age stem from the higher variation among users reflected in the standard deviation: in both age-subsamples there are one or two outliers which had an average of over twenty comments per post. Also, it needs to be pointed out that the age of the people commenting on posts was not considered, so when speaking of the younger and older subgroup, the age factor only applies to the people who initiated the communication with their post, while comment contributions may stem from all age groups.

Besides post format and interaction via comments, it is also important to consider what people post on Facebook, i.e. what do they communicate about? In the style of Bolander and Locher's (2010) content analysis, but with contextually adapted and slightly broader categories, table 3 shows the results of a simple categorization of posts according to thematic fields.

Table 3: Content analysis of Facebook posts

	Fijians 18-20 years	Fijians >= 25 years	Indo-Fijians 18-20 years	Indo-Fijians >= 25 years
Family & Partner	7	17	10	9
Self	29	16	24	28
Friends & Events	5	6	5	0
Insider Info	18	4	16	0
Interests	9	6	10	3
Religion	6	19	2	6
Jokes	6	1	4	13
News	0	0	0	6
Politics	1	8	1	0
Gossip	3	4	2	1
Work & School	3	0	3	0
Quotes & Advice	6	13	14	16
Wishes	1	0	1	8
Questions & Requests	2	3	8	0
Web finds	4	3	0	10

According to the numbers in table 3, most posts fall under the category “self”, regardless of subgroup. This category includes pictures of oneself as well as personal opinions and feelings. Of course this reflects to some degree what Facebook wants from its users, as the prompt for the input field reads “What’s on your mind?”. More interesting, however, seems to be the surprisingly low-yielding category “friends & events”, which really needs to be analyzed in combination with the category “insider info”. The latter contains anything that could not be made sense of by an outsider, i.e. posts that necessitate insider knowledge of a situation and therefore usually address one or a few specific friends. In that light it becomes apparent that the younger members of the community use the Facebook wall much more frequently for in-group interaction. Other notable observations are the high numbers in the “family & partner” category, especially among the older Fijian subgroup, consisting mainly of posted family pictures and a handful of declarations of affection. Further, the younger subgroups post more about their interests (hobbies, music, sports, etc.), whereas religion is a more prominent topic among the older groups. Posting quotes and advice (i.e. words of wisdom) is an overall popular content. The lack of target specific posts (insider info) among the older subgroups appears to be compensated by more general posts that also are less likely to evoke interaction beyond likes, i.e. jokes, religious comments (often praises of God), various findings from the web, and news.

From the findings presented in tables 1 – 3, it can be concluded that younger Facebook users use their wall more interactively, i.e. they produce more written posts that more often address their friends and regularly evoke more comments. The older Facebook users on the other hand seem to post more for the sake of sharing and interaction seems to be more limited to individual comments rather than unfolding dialogues.

5.2 How language is used on Facebook

Next to be addressed is the question which languages people use for Facebook communication in Fiji. With Fiji being a multilingual society, this question is of particular interest. For the analysis I distinguished between the language(s) of the post and the language(s) in the comment section. Whereas monolingual posts and comments are easy to code, ones that show elements of two languages are more difficult to categorize. Given the exploratory character of this study, I avoid discussing conceptual differences between code-switching, code-mixing and multilingualism and opted for a very simplistic categorization system, only making four distinction: all English, all Fijian or Hindi[5], and mixed with a distinction of predominant or first word language. This means that in the very rare cases where both languages were used in more or less equal parts, the first word was used for predominance categorization in order to avoid an additional category with little extra insights. The results are shown in table 4. Pictures containing language were not considered, i.e. only user-produced language, as for example memes taken from the internet have a strong English language bias.

Table 4: Languages used on Facebook wall posts and comments

	Fijians 18-20 years		Fijians >= 25 years		Indo-Fijians 18-20 years		Indo-Fijians >= 25 years	
	Posts	*Comm.*	*Posts*	*Comm.*	*Posts*	*Comm.*	*Posts*	*Comm.*
English	87.2%	62.8%	83.3%	35.7%	91.5%	77.7%	76.4%	58.1%
F / H	2.1%	30.7%	4.2%	31.5%	4.9%	8.7%	8.0%	14%
< E	7.2%	3.1%	10.2%	0.0%	1.0%	4.9%	2.9%	27.0%
< F / H	3.3%	3.2%	2.2%	32.6%	2.5%	7.3%	12.4%	0.7%
N: total	96	75	69	34	67	147	52	64

F = Fijian, H = Hindi, E = English, < E = mixed but predominantly English, < F/H = mixed but predominantly Fijian/Hindi, N = total raw numbers per subgroup

Not very surprisingly, English is the predominate language overall. However, there is an extremely interesting divide among subgroups on the one hand and Facebook context on the other. First of all, it can be noticed that posts differ quite a bit from comments. Whereas wall posts show much higher percentages of English-only use than comments across the board, Fijian- or Hindi-only usage is conversely much higher for comments than posts. This big difference suggests indeed that posts and comments serve two very different communicative functions on Facebook. The most likely interpretation seems to be that posts are a less directed form of communication for which English is more suitable, whereas comments often are direct exchanges with a specific person and therefore often carried out in the mother tongue. Similar observations have been made by Johnson (2013) with English-Welsh bilinguals on twitter. He interprets their increased use of Welsh in directed tweets as a form of Bell's (1984) audience design. Mixed

[5] There is a difference between Fiji-Hindi, the locally spoken variety, and (standard) Hindi, the standard (and usually written) variety. As I am not in the position to distinguish between the two and the focus here lies solely on the analysis of English, I use the label Hindi generically as cover term.

language communication is much less frequent, but among Fijians similar tendencies are visible, i.e. sentences tend to be more English with parts in Fijian in posts and vice versa in comments. The fact that these tendencies do not carry over to the Indo-Fijian subgroup is most likely due to the underlying low counts. The second major observation is that the younger groups have higher percentages of English-only in both posts and comments than the older subgroups, less Fijian- or Hindi-only use, and also slightly less mixed use overall. This may indeed suggest that English is endorsed more by the younger generation – at least in the CMC context – an observation that would confirm Tent's (2009) past finding that "the younger the respondent, the more positive the attitude toward English and the more it is used" (12).

Last but not least, all posts and comments where manually scanned for non-standard tokens which are typical of (English) language production in a CMC context. The chosen framework for categorization is adopted from Androutsopoulos (2007a) and consists of four categories, which in combination "lead to writing styles that are typical, if not even stereotypical, of informal online communication" (84, my translation). They are "Versprechsprachlichung" (orality), "mimisch-kinesische Kompensation" (mimic-kinetic compensation), "Ökonomisierung" (economy), and "Graphostilistik" (grapho-stylistics). Given a context of application where varying levels of language proficiency are the reality, it cannot always be clearly distinguished between an advertent manipulation of language and what may potentially be an inadvertent mistake or local language use. The first category is too wide for detailed application here, as it covers all aspects of language from syntax, discourse organization, lexis and even aspects of prosody, among others, largely following Koch and Oesterreicher's (1985) well-known model. But a typical example from the data that falls into this category, and probably the only one going by non-standard spelling only, is the written final consonant-cluster reduction *-n'* for *-ng*, a common feature in spoken English. Mimic-kinetic compensation includes features like verbal glosses or expressions of laughing (e.g. *lol*), but also emoticons, which I did not consider here. Economy comprises all time- and space-saving strategies, such as abbreviations. A nice example from the data is *grg* for *grog*, which is a local drink. Grapho-stylistics, finally, refers to the "manipulation of visual representation of language" (Androutsopoulos 2007a, 83, my translation), of which homophonic grapheme substitution is a typical example (e.g. numbers for letters). The following examples from the data will be discussed in more detail, as they show strong local influences. Although categorized as grapho-stylistics, the distinction from other categories can sometimes be blurred. The complete list of tokens can be found in the Appendix.

(1) *eht* (*it*), *iht* (*it*), *rili* (*really*), *bin* (*being*), *ich* (*each*), *kip* (*keep*), *swithrt* (*sweetheart*), *mit*/*-ing* (*meet*/*-ing*)
(2) *kudnt* (*couldn't*), *cudnt* (*couldn't*), *kul* (*cool*), *lukin* (*looking*), *schul* (*school*), *shud* (*should*), *cuk* (*cook*)
(3) *deh* (*the*), *d* (*the*), *dis* (*this*), *wid* (*with*), *den* (*then*/*than*), *doce* (*those*), *dey* (*they*), der (there)
(4) *aca* (*other*)
(5) *tew* (*too*)
(6) *lyk* (*like*), *tym* (*time*), *fyt* (*fight*), *fyn* (*fine*), *lyf* (*life*)
(7) *4rm* (*from*)

The following analysis is based on the description of Fiji English phonology in Tent and Mugler (2008). Examples (1) and (2) are most likely results of vowel length and quality neutralization, as monophthongs tend to be tense and short. Especially the tense quality appears to be expressed by the inserted *h* in *eht*/*iht*. KIT and FLEECE are not phonemically distinguished, although there is a small phonetic difference. The same is true for FOOT and GOOSE, hence probably the same graphic realization for both, *iht* and *kip*, as well as *kul* and *cuk*. The choice of the letter *u* for close back vowel representation may be a direct transfer from Fijian. The feature in (3) is certainly not exclusively local, but just as many non-standard varieties of English, Fiji English tends to have stops for the dental fricatives. A particularly nice example is *d*, which can therefore replace the entire article *the* on its own, just like single letters commonly replace entire words in CMC, such as *u* for *you*, *c* for *see*, *y* for *why*, etc. *Doce* in (3) is at the same time an example of frequent devoicing of syllable final /z/. Example (4) is as local as it gets. The spelling is a direct transfer from Fijian where *c* is the written representation of the dental fricative. In addition, influence non-rhotic pronunciation is found in the spelling of both this particular example and in other unmentioned ones as well. The spelling in (5) could simply be a creative use of a different homophonous spelling, but it probably only works because of the common yod-absence in /Cju/ syllables. The spelling representation of the PRICE diphthong as *y* in (6) does not appear to have a motivation rooted in L1, but it was found among different users. Finally, example (7) is a very interesting case, as the number 4 in combination with the letters *r* and *m* yield the word *form*, rather than *from*, which is a case of metathesis. Although metathesis is only documented for *ask* in Tent and Mugler (2008), the tendency for epenthetic vowel insertion in Fiji English clearly suggests an inclination to avoid (certain) consonant clusters.

Given that what is called Fiji English in the literature is usually described as a colloquially spoken variety, used for informal communication, it can be expected that some of the common features should be found in CMC as well, also in the light of the orality category above. Indeed, CMC has been suggested to be a relevant context for Fiji English usage by Mangubhai and Mugler (2003), as well as Zipp (2014). Most of the features found in Mugler and Tent's (2008) morphosyntactic description of Fiji English could be observed in the data. This not only confirms the use of (colloquial) Fiji English on Facebook (as opposed to standard (Fiji) English with fewer nativized features), but also opens up new perspectives for further research. If the features described in the literature are found on Facebook and thus retrievable via Graph Search, then it will be possible to analyze features of interest across various contexts for differences in constraints, either language specific for different modes or purely Facebook-based for global features of non-standard Englishes.

6. Conclusion

It has been the purpose of this paper to introduce Facebook Graph Search as a useful query tool for linguistic research. Graph Search can be used to easily retrieve language

samples of almost any language written[6] by people with access to Facebook with great precision. These language samples may be used for any type of explorative linguistic research and is not limited to sociolinguistics as long as the participants' interests are protected. Therefore it is imperative for researchers to adhere to common standards of research ethics when working with Graph Search, which may be especially limiting if examples of the data need to be published. Therefore, Graph Search is clearly not suitable for just any research project. Also, the quickly changing and often non-transparent character of such tools calls for critical and careful evaluation of the current suitability of Graph Search for any future project.

In the second part of this paper, a pilot study for such research was presented. The literature on language use in the CMC context in Fiji is not sufficient for gaining any first insights into the field, which makes it the perfect example for exploratory research with Graph Search. 400 wall posts and additional comments from 40 Facebook users were analyzed. Results showed that younger users apparently produce more text and more content directly addressed at Facebook friends, whereas older users opt more often for sharing content with unspecified recipients. A rough analysis of language use yielded patterns along the same lines, but also a noticeable difference between wall posts and comments: younger users showed more preference for English over Fijian or Hindi overall, but all users did so in wall posts compared to comments. The latter are likely understood to be more conversational and directly targeted at specific addressees and therefore evoke more instances of mother tongue language use. Last but not least, a qualitative look at the data for the occurrence of features typical of CMC as well as language features typical of Fiji English rounded off the pilot study. A multitude of features could be identified even in this relatively small sample of data, which shows that language use on Facebook in Fiji is comparable to CMC language described in the literature, making use of the same processes and similar features, but also showing clear influences of the local conditions. Much more data will be necessary to corroborate the findings from this pilot study and only data from additional contexts other than Facebook will allow for a generalization thereof across CMC modes. But given that this pilot study was also meant to outline the possibilities and potential benefits of Facebook Graph Search, it will hopefully be an impetus for more Graph Search-based research as well.

Acknowledgement

I would like to thank Marianne Hundt for helpful feedback on an earlier draft and the anonymous reviewers for their critical input. Also, I would like to acknowledge the forty anonymous Facebook users who unknowingly have been observed and without whom this study would not have been possible. Although I am fully aware that not all Facebook users may be conscious of the fact that their profile is publicly accessible, especially those who do deserve credit for allowing insights into their lives and making research such as this possible. As negative headlines about privacy issues and surveillance

[6] "Written" in the sense that Facebook is currently only searchable for written text. Common CMC practices show that this already exceeds the traditionally written languages as people start writing traditionally spoken languages as well. In the future audio and video material may increase the search space to spoken language.

accumulate, I hope this research is understood as a bonafide attempt to make positive use of given structures and not as a call to encourage developments in a certain direction.

Sources

Androutsopoulos, Jannis. 2007a. "Neue Medien – neue Schriftlichkeit?". *Mitteilungen des Deutschen Germanistenverbandes* 1/07, 72-97.

Androutsopoulos, Jannis. 2007b. "Style online: Doing hip-hop on the German-speaking Web". In: Peter Auer (ed.). *Style and social identities: alternative approaches to linguistic heterogeneity*. Berlin, NY: de Gruyter. 279-317.

Androutsopoulos, Jannis. 2008. "Potentials and limitations of discourse-centered online ethnography". *Language@internet* 5. <http://www.languageatinternet.org/articles/2008/1610> (accessed 06.10.14).

Androutsopoulos, Jannis. 2013. "Online data collection". In: Christine Mallinson, Becky Childs, Gerard Van Herk (eds.). *Data Collection in Sociolinguistics: Methods and Applications*. London: Routledge. 236–249.

Bell, Allan. 1984. "Language style as audience design". *Language in Society* 13 (2). 145-204. DOI: 10.1017/S004740450001037X

Bolander, Brook and Miriam A. Locher. 2010. "Constructing Identity on Facebook." In Karen Junod and Didier Maillat (eds.). *Performing the Self. Swiss Papers in English Language and Literature* 24. Tübingen: Narr. 165–188.

Bolander, Brook and Miriam Locher. 2014. "Doing sociolinguistic research on computer-mediated data: A review of four methodological issues." *Discourse, Context and Media* 3. 14-26. DOI: 10.1016/j.dcm.2013.10.004

Cave, Danielle. 2012. "Digital Islands: How the Pacific's ICT Revolution Is Transforming the Region". Sydney: Lowy Institute for International Policy. <http://www.lowyinstitute.org/files/cave_digital_islands_web.pdf> (accessed 07.10.14)

Honeycutt, Courtenay and Daniel Cunliffe. 2010. "The Use of the Welsh Language on Facebook". *Information, Communication & Society* 13 (2). 226-248. DOI: 10.1080/13691180902914628

Johnson, Ian. 2013. "Audience Design and Communication Accommodation Theory: Use of Twitter by Welsh – English Biliterates". In: Elin Haf Gruffydd Jones and Enrique Uribe-Jongbloed (eds.). *Social Media and Minority Languages – Convergence and the Creative Industries*. Bristol: Multilingual Matters. 99-118.

Koch, Peter and Wulf Oesterreicher. 1985. "Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte". *Romanistisches Jahrbuch* 36, 15-43.

Landert, Daniela and Andreas H. Jucker. 2011. "Private and public in mass media communication: From letters to the editor to online commentaries". *Journal of Pragmatics* 43. 1422-1434.

Magubhai, Francis and France Mugler. 2003. "The Language Situation in Fiji". *Current Issues in Language Planning* 4:3-4, 367-459.

Mugler, France and Jan Tent. 2008. "Fiji English: morphology and syntax". In: Kate Burridge and Bernd Kortmann (eds.). *Varieties of English 3*. Berlin: Mouton de Gruyter. 546-567.

Solomon, Kate. 16.1.2013. “Facebook Graph Search: first impressions”. *Techradar.com* <http://www.techradar.com/news/internet/facebook-graph-search-first-impressions-1125472> (accessed 27.10.2014)
Steinfeld, Nili and Azi Lev-On. 2014. “‘Well-done, Mr. Mayor!’: Linguistic analysis of municipal facebook pages”. *Proceedings of the 15th Annual International Conference on Digital Government Research*. New York: ACM. 273-279. DOI: 10.1145/2612733.2612763
Sultana, Shaila. 2014. “Young adults’ linguistic manipulation of English in Bangla in Bangladesh”. *International Journal of Bilingual Education and Bilingualism* 17 (1). 74-89. DOI: 10.1080/13670050.2012.738644
Tent, Jan and France Mugler. 2008. "Fiji English: phonology". In: Kate Burridge and Bernd Kortmann (eds.). *Varieties of English 3*. Berlin: Mouton de Gruyter. 234-266.
Tent, Jan. 2009. “Language Use and Attitudes in Fiji 1993–2005”. Paper given at the University of Zurich, Switzerland on 26/05/09 [handout].
West, Laura E. 2013. “Facebook sharing: A sociolinguistic analysis of computer-mediated storytelling”. *Discourse, Context & Media* 2 (1). 1-13.
Wilson, Christo, Bryce Boe, Alessandra Sala, Krishna P.N. Puttaswamy, and Ben Y. Zhao. 2009. “User interactions in social networks and their implications”. *Proceedings of the 4th ACM European conference on Computer systems* (EuroSys '09). New York: ACM. 205-218. DOI=10.1145/1519065.1519089
Zipp, Lena. 2014. "Indo-Fijian English - linguistic diaspora or endonormative stabilization?" In: Marianne Hundt and Devyani Sharma (eds.). *English in the Indian Diaspora*. Amsterdam: John Benjamins [Varieties of English Around the World]. 187-213.

Appendix

Orality
havin, payin, performin, shopin, movin, greetin, disturbin, thinkin, startin, darlin

Mimic-kinetic Compensation
– all capital letters
– capital letters for stress (*go out and WIN*)
– lengthening for stress (*plizzzz, okaay*)
lmaooo, lol, xx, xxx, lolz, lols, lm(f)ao

Economy
– all small letters
– no apostrophes
– no spaces

bro = brother	*grg = grog*	*jst =just*
pm = private message	*bday = birthday*	*dnt = don’t*
hvg =having	*b’day = birthday*	*knw = know*
nw = now	*self = myself*	*frm = from*

i k r = I know right
bt = but
fb = facebook
nt = not
gt = got
cnt = can't
hw = how
dt = that
hr = her
hm = home
mustv = must've
k = ok
thnx = thanks
daxz = thanks
Thks = thanks
thanxxxsh = thanks
nxt = next
abt = about
hme = home
hav = have
tht = that
GB = god bless
FML = fuck my life
TGIF = thank god it's Friday
grl = girl
esp = especially
fav = favorite
fes = festival
I.L.Y = I love you
frnd(s) = friend(s)
hve = have
Thurs = Thursday
Gudmrng = good morning
Goodmrng = good morning
cm(e) = come
wedg = wedding
sry = sorry
nva = never
thn = then
adjst = adjust
txt = text
msgs = messages
wnt = want
pic = picture
bsd = beside
othr = other
plce = place
TBH = to be honest
d.p = display picture
LMS = like my status
mke = make
pple = people
aniwys =anyways
ceremny = ceremony
tho(u) = though

Grapho-stylistics

u = you
y = why
c = see
v = we
d = the
b = be
m = I'm
a = are
n = and
nd = and
ur = your
u'r = you're
urself = yourself
luv = love
sista = sister
cuz = cousin
cuzn = cousin
2 = to
2moro = tomorrow
2day = today
2dae = today
@ = at
4 = for
b4 = before
4rm = from
gr8 = great
ani1 = anyone
every1 = everyone
gal = girl
coz = because
eht = it
iht = it
itz = it's
yah = you
yhu = you
yuh = you
youuh = you
yo = your
baybeh = baby
baybaaaay = baby
daii = day
CH33RZ = cheers
deh = the
da = the
kruw = crew
mah = my
ma = my
songz = songs
h8t = hate
aca = other
kul = cool
tew = too
bak = back
lukin = looking
cuk = cook
schul = school
kudnt = couldn't
(al)rite = (al)right
nite = night
tonite = tonight
luh = love
mwnin = morning
chix = chicks
thang = thing
waz = was
woz = was
mahn = man
Redz = Reds
hia = here
lyk = like
tym = time
tymz = times
fyt = fight
tym = time
fyn = fine

lyf = life
pliz = please
datz = that's
eatz = eats
itz = it's
diz = this
thnkz = thanks
pliz = please
dunno = don't know
Imma = I'm gonna
biq = big
lonq = long
leqendary = legendary
qet = get
cauqht = caught
cominq = coming
dayum = damn
afta = after
sum = some
phocking = fucking
fxck = fuck

shyt = shit
dat = that
wat = what
dis = this
wid = with
everyfngs = everything's
shud = should
jaz = just
rili = really
bin = being
ich = each
kip = keep
swithrt = sweetheart
kidin = kidding
beta = better
of coz = of course
doce = those
dey = they
den = than
der = there
muj = much

hu = who
cum = come
ol = all
wij you = with you
verii =very
gud = good
motherfuqker = motherfucker
fuqk = fuck
muhfuckas = motherfuckers
cudnt = couldn't
wea = where
dea = their
olways = always
havn't = haven't
dint = didn't
mit(ing) = meet(ing)

Potential Misspellings

teaches = Teachers
al = all
shal = shall
I'l = I'll
wel = well
tel = tell
expell =expel
sik(ness) = sick(ness)

seriousli = seriously
realy = really
reali = really
belives = believes
than = then
planty = plenty
awesum = awesome
disapper = disappear

dam = damn
of = off
wishpers = whispers
hapend = happened
wana = wanna
toked = talked
wateva =whatever
wen = when

Bio-Note

André Huber is a research and teaching assistant at the English Department of the University of Zürich. Currently, his research focuses on various aspects of the new media and their influence on English in Fiji. His research interests include Sociolinguistics, CMC and Corpus Linguistics.